

ABSTRACT OF THE DISCLOSURE

The present invention provides a facility for discovering a set of inference rules, such as “*X is author of Y \approx X wrote Y*”, “*X solved Y \approx X found a solution to Y*”, and “*X caused Y \approx Y is triggered by X*”, by analyzing a corpus of natural language text. The corpus is parsed to identify grammatical relationships between words and to build dependency trees formed of the relationships between the words. Paths linking words in the dependency trees are identified. If two paths tend to link the same sets of words, their meanings are taken to be similar. An inference rule is generated for each pair of similar paths. The output of the inventive system is a set of inference rules and a database in which to store these inference rules. The rules generated by the system are interpretable by machines and used in other applications (e.g. information extraction, information retrieval, and machine translation).